

Description

Isolation of the biosynthesis genes for pseudo-oligosaccharides from *Streptomyces glaucescens* GLA.O, and their use

5 The present invention relates to the isolation of genes which encode enzymes for the biosynthesis of α -amylase inhibitors, so-called pseudo-oligosaccharides. The genes concerned are, in particular, genes from the Streptomyces strain *Streptomyces glaucescens* GLA.O (DSM 40716). In
10 addition, this present patent describes the use of these genes for producing acarbose and homologous substances with the aid of *Streptomyces glaucescens* GLA.O, the heterologous expression of these genes in other strains which produce pseudo-oligosaccharides (e.g. *Actinoplanes* sp SE50/100) for the purpose of increasing and stabilizing
15 production, and also their heterologous expression in other microorganisms such as *E. coli*, *Bacillus subtilis*, Actinomycetales, such as *Streptomyces*, *Actinoplanes*, *Ampullariella* and *Streptoporangium* strains, *Streptomyces hygroscopicus* var. *limoneus* and *Streptomyces glaucescens*, and also biotechnologically relevant fungi (e.g. *Aspergillus niger* and *Penicillium chrysogenum*) and yeasts (e.g. *Saccharomyces cerevisiae*). The invention also relates to homologous genes in other
20 microorganisms and to methods for isolating them.

Streptomyces glaucescens GLA.O produces the two antibiotics
25 hydroxystreptomycin (Hütter (1967) Systematik der Streptomyceten (Taxonomy of the Streptomycetes). Basel, Karger Verlag) and tetracenomycin (Weber et al. (1979) Arch. Microbiol. 121: 111-116). It is known that streptomycetes are able to synthesize structurally varied natural products. However, the conditions under which these compounds are
30 produced are frequently unknown, or else the substances are only produced in very small quantities and not detected.

The α -amylase inhibitor acarbose has been isolated from a variety of *Actinoplanes* strains (SE50, SE82 and SE18) (Schmidt et al. (1977)
35 Naturwissenschaften 64: 535-536). This active substance was discovered in association with screening for α -amylase inhibitors from organisms of the genera *Actinoplanes*, *Ampullariella* and *Streptosporangium*. Acarbose is pseudotetrasaccharide which is composed of an unusual unsaturated

cyclitol unit to which an amino sugar, i.e. 4,6-dideoxy-4-amino-D-glucopyranose, is bonded. Additional α -1,4-glycosidically linked D-glucopyranose units can be bonded to the amino sugar. Thus, acarbose, for example, contains two further molecules of D-glucose. The producing strain synthesizes a mixture of pseudo-oligosaccharide products which possess sugar side chains of different lengths (Schmidt et al. (1977) Naturwissenschaften 64: 535-536). The acarbose cyclitol residue is identical to the compound valienamine, which is a component of the antibiotic validamycin A (Iwasa et al. (1979) J. Antibiot. 32: 595-602) from *Streptomyces hygroscopicus* var. *limoneus*.

Acarbose can be produced by fermentation using an *Actinoplanes* strain and has achieved great economic importance as a therapeutic agent for diabetics. While *Actinoplanes* synthesizes a mixture of α -amylase inhibitor products, it is only the compound having the relative molecular weight of 645.5 (acarviosin containing 2 glucose units (Truscheit (1984) VIIIth International Symposium on Medicinal Chemistry, Proc. Vol. 1. Swedish Academy of Pharmaceutical Sciences, Stockholm, Sweden), which is employed under the generic name of acarbose. The fermentation conditions are selected to ensure that acarbose is the main product of the fermentation. Alternatives are to use particular selectants and strains in which acarbose is formed as the main product or to employ purification processes for achieving selective isolation (Truscheit (1984) VIIIth International Symposium on Medicinal Chemistry, Proc. Vol. 1. Swedish Academy of Pharmaceutical Sciences, Stockholm, Sweden). It is also possible to transform the product mixture chemically in order, finally, to obtain the desired product acarbose.

In contrast to the genus *Streptomyces*, the genus *Actinoplanes* has not so far been investigated intensively from the genetic point of view. Methods which were established for the genus *Streptomyces* are not transferable, or are not always transferable, to the genus *Actinoplanes*. In order to use molecular biological methods to optimize acarbose production in a purposeful manner, the genes for acarbose biosynthesis have to be isolated and characterized. In this context, the possibility suggests itself of attempting to set up a host/vector system for *Actinoplanes* sp. However, this is very tedious and elaborate owing to the fact that studies on *Actinoplanes* have been relatively superficial.

The invention described in the present patent application achieves the object of cloning the biosynthesis genes for acarbose and homologous pseudo-oligosaccharides, with these genes being cloned from
5 Streptomyces glaucescens GLA.O, which is a streptomycete which has been thoroughly investigated genetically (Crameri et al. (1983) J. Gen. Microbiol. 129: 519-527; Hintermann et al. (1984) Mol. Gen. Genet. 196: 513-520; Motamedi and Hutchinson (1987) PNAS USA 84: 4445-4449; Geistlich et al. (1989) Mol. Microbiol. 3: 1061-1069) and which, surprisingly,
10 is an acarbose producer. In starch-containing medium, Streptomyces glaucescens GLA.O produces pseudo-oligosaccharides having the molecular weights 645, 807 and 970.

Part of the subject matter of the invention is, therefore, the isolation of the
15 corresponding biosynthesis genes from Streptomyces glaucescens GLA.O and their use for isolating the adjoining DNA regions in order to complete the gene cluster of said biosynthesis genes.

The isolation of the genes for biosynthesizing pseudo-oligosaccharides,
20 and the characterization of these genes, are of great importance for achieving a better understanding of the biosynthesis of the pseudo-oligosaccharides and its regulation. This knowledge can then be used to increase the productivity of the Streptomyces glaucescens GLA.O strain with regard to acarbose production by means of established classical and
25 molecular biological methods. In addition to this, the entire gene cluster which encodes the synthesis of the pseudo-oligosaccharides, or individual genes from this gene cluster, can also be expressed in other biotechnologically relevant microorganisms in order to achieve a further increase in, or a simplification of, the preparation of pseudo-
30 oligosaccharides such as acarbose. Specific modification of the biosynthesis genes can also be used to prepare a strain which exclusively produces acarbose having a molecular weight of 645. Since the genes for biosynthesizing antibiotics are always present in clusters and are often very strongly conserved (Stockmann and Piepersberg (1992) FEMS Microbiol.
35 Letters 90: 185-190; Malpartida et al. (1987) Nature 314:642-644), the Streptomyces glaucescens GLA.O genes can also be used as a probe for isolating the acarbose-encoding genes from Actinoplanes sp., for example. The expression of regulatory genes, or of genes which encode limiting

steps in the biosynthesis, can result in productivity in *Streptomyces glaucescens* GLA.O, *Actinoplanes* sp. or corresponding producer strains being increased. An increase in productivity can also be achieved by switching off (knocking out or mutagenizing) those acarbose biosynthesis genes which have an inhibitory effect in the biosynthesis.

One possible strategy for cloning antibiotic biosynthesis genes which have not previously been isolated is that of using gene-specific probes (Stockmann and Piepersberg (1992) FEMS Microbiol. Letters 90: 185-190; 10 Malpartida et al. (1987) Nature 314:642-644). These probes can be DNA fragments which are P³²-labeled or labeled in some other way; otherwise, the appropriate genes can be amplified directly from the strains to be investigated using degenerate PCR primers and isolated chromosomal DNA as the template.

15 The latter method has been employed in the present study. Pseudo-oligosaccharides such as acarbose contain a 4,6-deoxyglucose building block as a structural element. The enzyme dTDP-glucose 4,6-dehydratase is known to be involved in the biosynthesis of 4,6-deoxyglucose 20 (Stockmann and Piepersberg (1992) FEMS Microbiol, Letters 90: 185-190). Since deoxysugars are a frequent constituent of natural products and antibiotics, this enzyme may possibly be a means for isolating the corresponding antibiotic biosynthesis genes. Since these genes are always present as clusters, it is sufficient to initially isolate one gene; the isolation 25 and characterization of the adjoining DNA regions can then be undertaken subsequently.

For example a dTDP-glucose 4,6-dehydratase catalyzes a step in the biosynthesis of hydroxystreptomycin in *Streptomyces glaucescens* GLA.O 30 (Retzlaff et al. (1993) Industrial Microorganisms. Basic and applied molecular genetics ASM, Washington DC, USA). Further dTDP-glucose 4,6-dehydratases have been isolated from other microorganisms, for example from *Streptomyces griseus* (Pissowotzki et al. (1991) Mol. Gen. Genet. 231: 113-123), *Streptomyces fradiae* (Merson-Davies and 35 Cundcliffe (1994) Mol. Microbiol. 13: 349-355) and *Streptomyces violaceoruber* (Bechthold, et al. (1995) Mol. Gen. Genet. 248: 610-620).

It was consequently possible to deduce the sequences for the PCR primers for amplifying a dTDP-glucose 4,6-dehydratase from the amino acid sequences of already known biosynthesis genes. For this, conserved regions in the protein sequences of these enzymes were selected and the amino acid sequences were translated into a nucleic acid sequence in accordance with the genetic code. The protein sequences were taken from the EMBL and Genbank databases. The following sequences were used: Streptomyces griseus; accession number: X62567 gene: strE (dated 10.30.1993); Streptomyces violaceoruber; accession number: L37334 gene: graE (dated 04.10.1995); Saccharopolyspora etythraea; accession number: L37354 gene: gdh (dated 11.09.1994). A large number of possible primer sequences are obtained as a result of the degeneracy of the genetic code. The fact that streptomycetes usually contain a G or C in the third position of a codon (Wright and Bibb (1992) gene 113: 55-65) reduces the number of primers to be synthesized. These primer mixtures can then be used to carry out a PCR amplification with the DNA from strains to be investigated, with the amplification ideally leading to an amplified DNA fragment. In the case of highly conserved proteins, this fragment is of a predictable length which ensues from the distance between the primers in the nucleic acid sequence of the corresponding gene. However, an experimental mixture of this nature does not inevitably have to result in an amplificate. The primers may be too unspecific and amplify a very large number of fragments; alternatively, no PCR product is obtained if there are no complementary binding sites in the chromosome for the PCR primers which have been prepared.

The investigation of the streptomycete strain Streptomyces glaucescens GLA.O resulted in an amplified DNA fragment (acbD) which had the expected length of 550 bp. Further investigation showed that, besides containing a dTDP-glucose 4,6-dehydratase gene for biosynthesizing hydroxystreptomycin, this strain surprisingly contains a second dTDP-glucose 4,6-dehydratase gene for biosynthesizing pseudo-oligosaccharides such as acarbose. While the two genes exhibit a high degree of homology, they are only 65% identical at the amino acid level.

The acbD probe (see Example 2 and Table 2A) was used to isolate, from Streptomyces glaucescens GLA.O, a 6.8 kb PstI DNA fragment which

encodes a variety of genes (acbA, acbB, acdC, acbD, acbE and acbF) which are involved in the biosynthesis of the pseudo-oligosaccharides.

Deleting the acbBCD genes (aminotransferase, acbB, dTDP-glucose
5 synthase, acbC, dTDP-glucose 4,6-dehydratase, acbD, see Example 6)
resulted in the production of a mutant of *Streptomyces glaucescens* GLA.O
which no longer produces any pseudo-oligosaccharides in the production
medium. The involvement of the acbBCD genes in the synthesis of
pseudo-oligosaccharides was therefore verified by deleting the
10 corresponding loci.

The two genes, i.e. dTDP-glucose synthase and dTDP-glucose 4,6-
dehydratase, ought to be involved in the biosynthesis of the deoxysugar of
the pseudo-oligosaccharides, as can be concluded from the function of
15 thoroughly investigated homologous enzymes (see above). The amino-
transferase (encoded by the acbB gene) is probably responsible for
transferring the amino group either to the sugar residue or to the cyclitol
residue. By analyzing the protein sequence of acbB, an amino acid motif
was found which is involved in binding pyridoxal phosphate. This motif is
20 typical of class III aminotransferases (EC 2.6.1.11; EC 2.6.1.13; EC
2.6.1.18; EC 2.6.1.19; EC 2.6.1.62; EC 2.6.1.64; EC 5.4.3.8). The precise
enzymic function of acbB can only be elucidated by further investigation of
the biosynthesis of the pseudo-oligosaccharides. acbE encodes a
transcription-regulating protein which exhibits a great deal of similarity to
25 DNA-binding proteins which possess a helix-turn-helix motif (e.g. *Bacillus*
subtilis DegA, P37947; Swiss-Prot database). Thus, the transcription
activator CcpA from *Bacillus subtilis* inhibits the formation of α -amylase in
the presence of glucose, for example (Henkin et al. (1991) *Mol. Microbiol.*
5: 575-584). Other representatives of this group are proteins which
30 recognize particular sugar building blocks and are able to exhibit a positive
or negative effect on the biosynthesis of metabolic pathways. The
biosynthesis of the pseudo-oligosaccharides is also regulated in
Streptomyces glaucescens GLA.O. It was only previously possible to
demonstrate the synthesis of pseudo-oligosaccharides on starch-
35 containing media. While this method indicated that AcbE might be
responsible for regulating pseudo-oligosaccharide synthesis, the precise
mechanism is still not known. However, molecular biological methods can
now be used to modify the gene specifically in order to obtain an increased

rate of pseudo-oligosaccharide biosynthesis. Furthermore, the DNA site at which *acbE* binds can be identified by means of so-called gel shift assays (Miwa et al. (1994) Microbiology 140: 2576-2575). An increase in the rate at which acarbose is biosynthesized can be achieved after identifying and
5 then modifying the promoters and other regulatory DNA regions which are responsible for the transcription of the pseudo-oligosaccharide genes.

At present, the function of *acbF* is still not definitely known. The corresponding gene product exhibits homologies with sugar-binding
10 proteins such as the sugar-binding protein from *Streptococcus mutans* (MsmE; Q00749: Swissprot database), making it probable that it is involved in the biosynthesis of the pseudo-oligosaccharides. The gene product of the *acbA* gene exhibits homologies with known bacterial ATP-binding proteins (e.g. from *Streptomyces peucetis* DrrA, P32010: SwissProt
15 database). The *AcbA* protein possesses the typical ATP/GTP binding motif, i.e. the so-called P loop. These proteins constitute an important component of so-called ABC transporters, which are involved in the active transport of metabolites at biological membranes (Higgins (1995) Cell 82: 693-696). Accordingly, *AcbA* could be responsible for exporting pseudo-
20 oligosaccharides out of the cell or be involved in importing sugar building blocks for biosynthesizing α -amylase inhibitors such as maltose.

All streptomycete genes for biosynthesizing secondary metabolites which have so far been analyzed are arranged in a cluster. For this reason, it is to
25 be assumed that the acarbose biosynthesis genes according to the application are also arranged in such a gene cluster. The remaining genes which are relevant for pseudo-oligosaccharide biosynthesis can therefore also be isolated by isolating the DNA regions which adjoin the 6.8 kb *Pst*I DNA fragment according to the invention. As has also already been
30 mentioned above, it is readily possible to isolate homologous gene clusters from microorganisms other than *Streptomyces glaucescens* GLA.O.

The invention therefore relates to a recombinant DNA molecule which comprises genes for biosynthesizing acarbose and homologous pseudo-
35 oligosaccharides, in particular a recombinant DNA molecule in which individual genes are arranged, with respect to their direction of transcription and order, as depicted in Figure 3 and/or which exhibits a restriction

enzyme cleavage site pattern as depicted in Figure 3, and, preferably, to a recombinant DNA molecule which

- (a) comprises a DNA sequence according to Table 4, or parts thereof;
- (b) comprises a DNA sequence which is able to hybridize, under
5 stringent conditions, with the DNA molecule according to (a), or parts thereof; or
- (c) comprises a DNA sequence which, because of the degeneracy of the genetic code, differs from the DNA molecules according to (a) and (b) but which permits the expression of the proteins which can
10 be correspondingly expressed using the DNA molecule according to (a) and (b), or parts thereof.

The invention furthermore relates to a recombinant DNA molecule which comprises the *acbA* gene, in particular which is characterized in that it
15 comprises the DNA sequence of nucleotides 1 to 720 according to Table 4, or parts thereof; to a recombinant DNA molecule which comprises the *acbB* gene, in particular which is characterized in that it comprises the DNA sequence of nucleotides 720 to 2006 according to Table 4, or parts thereof; to a recombinant DNA molecule which comprises the *acbC* gene,
20 in particular which is characterized in that it comprises the DNA sequence of nucleotides 2268 to 3332 according to Table 4, or parts thereof; to a recombinant DNA molecule which comprises the *acbD* gene, in particular which is characterized in that it comprises the DNA sequence of nucleotides 3332 to 4306 according to Table 4, or parts thereof; to a
25 recombinant DNA molecule which comprises the *acbE* gene, in particular which is characterized in that it comprises the DNA sequence of nucleotides 4380 to 5414 according to Table 4, or parts thereof; and to a recombinant DNA molecule which comprises the *acbF* gene, in particular which is characterized in that it comprises the DNA sequence of
30 nucleotides 5676 to 6854 according to Table 4, or parts thereof.

The invention furthermore relates to oligonucleotide primers for the PCR amplification of a recombinant DNA molecule which is as described above and which comprises genes for biosynthesizing acarbose and homologous
35 pseudo-oligosaccharides, with the primers having, in particular, the sequence according to Table 1.

The invention furthermore relates to a vector which comprises a recombinant DNA molecule which comprises a DNA molecule as described in the penultimate and prepenultimate paragraphs, in particular which is characterized in that the vector is an expression vector and said DNA molecule is linked operatively to a promoter sequence, with the vector preferably being suitable for expression in host organisms which are selected from the group consisting of *E. coli*, *Bacillus subtilis*, Actinomycetales, such as *Streptomyces*, *Actinoplanes*, *Ampullariella* and *Streptosporangium* strains, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens* and also biotechnologically relevant fungi (e.g. *Aspergillus niger*, *Penicillium chrysogenum*) and yeasts (e.g. *Saccharomyces cerevisiae*), with *Streptomyces glaucescens* GLA.O or *Actinoplanes* sp. being very particularly preferred. Since the operative linkage of said DNA molecule to promoter sequences of the vector is only one preferably embodiment of the invention, it is also possible for expression to be achieved using promoter sequences which are endogenous in relation to the DNA molecule, e.g. the promoters which are in each case natural, or the natural promoters which have been mutated with regard to optimizing the acarbose yield. Such natural promoters are part of the DNA molecule according to the invention.

The invention furthermore relates to a vector which comprises a DNA molecule according to the invention for use in a process for eliminating or altering natural acarbose biosynthesis genes in an acarbose-producing microorganism. Such a vector is preferably selected from the group consisting of pGM160 and vectors as described in European Patents EP 0 334 282 and EP 0 158 872.

The invention furthermore relates to a host cell which is transformed with one of the above-described DNA molecules or vectors, in particular characterized in that said host cell is selected from the group consisting of *E. coli*, *Bacillus subtilis*, Actinomycetales, such as *Streptomyces*, *Actinoplanes*, *Ampullariella* or *Streptosporangium* strains, *Streptomyces hygroscopicus* var. *limoneus* or *Streptomyces glaucescens*, and also biotechnologically relevant fungi (e.g. *Aspergillus niger* and *Penicillium chrysogenum*) and yeasts (e.g. *Saccharomyces cerevisiae*); it is very particularly preferred for it to be selected from the group consisting of *Streptomyces glaucescens* GLA.O and *Actinoplanes* sp.

The invention furthermore relates to a protein mixture which can be obtained by expressing the genes of the recombinant DNA molecule according to the invention, comprising genes for biosynthesizing acarbose and homologous pseudo-oligosaccharides, in particular characterized in that the DNA molecule

- 5 (a) comprises a DNA sequence according to Table 4, or parts thereof;
- (b) comprises a DNA sequence which is able to hybridize, under stringent conditions, with the DNA molecule according to (a) or parts thereof; or
- 10 (c) comprises a DNA sequence which, because of the degeneracy of the genetic code, differs from the DNA molecules according to (a) and (b) but which permits the expression of the proteins which can correspondingly be expressed using the DNA molecule according to (a) and (b), or parts thereof.

15

The invention furthermore relates to isolated proteins which can be obtained by expressing the genes which are encoded by the DNA molecule described in the previous paragraph.

- 20 The following statements apply to all the individual genes identified within the context of the present invention and have only been brought together for reasons of clarity: the invention furthermore relates to a protein which is encoded by a recombinant DNA molecule as described in the last paragraph but one, in particular characterized in that it comprises the DNA
- 25 sequence of nucleotides 1 to 720 or 720 to 2006 or 2268 to 3332 or 3332 to 4306 or 4380 to 5414 or 5676 to 6854 according to Table 4 or parts thereof; a protein is very particularly preferred which is encoded by the acbA gene or the acbB gene or the acbC gene or the acbD gene or the acbE gene or the acbF gene, and which comprises the amino acid
- 30 sequence according to Table 4 or parts thereof.

The invention furthermore relates to a process for obtaining the proteins which were described above as being part of the subject-matter of the invention, which process is characterized in that

- 35 (a) the proteins are expressed in a suitable host cell, in particular which is characterized in that said host cell is selected from the group consisting of *E. coli*, *Bacillus subtilis*, Actinomycetales, such as *Streptomyces*, *Actinoplanes*, *Ampullariella* or *Streptosporangium*

- 5 strains, *Streptomyces*, *hygroscopicus* var. *limoneus* or *Streptomyces glaucescens*, and also biotechnologically relevant fungi (e.g. *Aspergillus niger* and *Penicillium chrysogenum*) and yeasts (e.g. *Saccharomyces cerevisiae*); with the host cell very particularly preferably being selected from the group consisting of *Streptomyces glaucescens* GLA.O and *Actinoplanes* sp., and
- (b) are isolated.

10 The invention furthermore relates to a process for preparing acarbose, characterized in that

- (a) one or more genes of the recombinant DNA molecule which comprises a DNA sequence according to Table 4 or parts thereof or which comprises a DNA sequence which is able to hybridize, under stringent conditions, with the DNA molecule according to Table 4, or
- 15 parts thereof, or which comprises a DNA sequence which, because of the degeneracy of the genetic code, differs from the DNA molecules which have just been described but which permits the expression of the proteins which can be correspondingly expressed using these DNA molecules, or parts thereof, is/are used for
- 20 expression in a suitable host cell which is selected, in particular, from the same group as in the last paragraph, and
- (b) the acarbose is isolated from culture supernatants of said host cell.

25 The invention furthermore relates to a process for preparing acarbose, characterized in that

- (a) one or more genes of the recombinant DNA molecule which comprises a DNA sequence according to Table 4 or parts thereof or which comprises a DNA sequence which is able to hybridize, under stringent conditions, with the DNA molecule according to Table 4, or
- 30 parts thereof, or which comprises a DNA sequence which, because of the degeneracy of the genetic code, differs from the DNA molecules which have just been described but which permits expression of the proteins which can be correspondingly expressed using the DNA molecules, or parts thereof, are eliminated in an
- 35 acarbose-producing host cell, in particular *Streptomyces glaucescens* GLA.O and *Actinoplanes* sp., and
- (b) the acarbose is isolated from said host cell.

In this connection, the elimination of one or more genes can be effected by means of standard molecular biological methods, for example using the above-described vectors (pGM160 and others). A gene to be eliminated could, for example, be the *acbE* gene, which probably has a regulatory function. Genes could likewise be eliminated with the aim of obtaining pure acarbose as the only fermentation product and no longer obtaining a mixture of homologous pseudo-oligosaccharides (see above). The elimination of said genes is preferably achieved using the vectors which have been described above for this purpose.

10

The invention furthermore relates to a process for preparing acarbose, characterized in that the processes for preparing acarbose which have been described in the previous two paragraphs are combined with each other, such that, therefore, one or more of said genes is/are expressed artificially and one or more of said genes is/are eliminated.

15

The invention furthermore relates to a process for altering the gene expression of endogenous acarbose biosynthesis genes by mutating the respective gene promoter in order to obtain improved yields of acarbose. In this context, known methods of homologous recombination can be used to introduce the mutations into the production strain to be improved. These mutations can be transitions, deletions and/or additions. An "addition" can, for example, denote the addition of one single nucleotide or several nucleotides or of one or more DNA sequences which have a positive regulatory effect and which bring about an enhancement of the expression of an endogenous gene for biosynthesizing acarbose. The converse case, i.e. the addition of a DNA sequence which has a negative regulatory effect for repressing an endogenous acarbose biosynthesis gene is also a preferred form of an addition. "Transitions" may, for example, be nucleotide exchanges which reduce or amplify the effect of regulatory elements which act negatively or positively. "Deletions" can be used to remove regulatory elements which act negatively or positively. The endogenous genes of this process are preferably present in Actinomycetales, such as *Streptomyces*, *Actinoplanes*, *Ampullariella* or *Streptosporangium* strains, *Streptomyces hygroscopicus* var. *limoneus* or *Streptomyces glaucescens*; very particularly, they are present in *Streptomyces glaucescens* GLA.O and *Actinoplanes* sp.

20

25

30

35

The invention furthermore relates to the use of *Streptomyces* GLA.O for obtaining acarbose.

5 The invention furthermore relates to the use of *Streptomyces* GLA.O for preparing mutants of this strain by the "classical route", which mutants make it possible to achieve a more abundant production of acarbose. The methods for preparing improved natural product producers of this nature have been known for a long time and frequently make use of classical steps of mutagenesis and selection.

10

The invention furthermore relates to a process for completing the gene cluster for biosynthesizing acarbose and homologous polysaccharides according to Table 4, characterized in that

- 15 a) hybridization probes which are derived from the DNA molecule according to Table 4 are prepared,
b) these hybridization probes are used for the genomic screening of DNA libraries obtained from *Streptomyces glaucescens* GLA.O, and
c) the clones which are found are isolated and characterized.

20 The invention furthermore relates to a process for completing the gene cluster for biosynthesizing acarbose and homologous pseudo-oligosaccharides according to Table 4, characterized in that, proceeding from the recombinant DNA molecule according to Table 4,

- a) PCR primers are prepared,
25 b) these PCR primers are used to accumulate DNA fragments of genomic DNA from *Streptomyces glaucescens* GLA.O, with these primers being combined with those primers which hybridize from sequences of the vector system employed,
c) the accumulated fragments are isolated and characterized.

30

The invention furthermore relates to a process for isolating a gene cluster for biosynthesizing acarbose and homologous pseudo-oligosaccharides from acarbose-producing microorganisms other than *Streptomyces glaucescens* GLA.O, characterized in that, proceeding from the recombinant DNA molecule according to Claim 4,

35

- a) hybridization probes are prepared,

- b) these hybridization probes are used for the genomic or cDNA screening of DNA libraries which have been obtained from the corresponding microorganism, and
- c) the clones which are found are isolated and characterized.

5

The invention furthermore relates to a process for isolating a gene cluster for biosynthesizing acarbose and homologous pseudo-oligosaccharides from acarbose-producing microorganisms other than *Streptomyces glaucescens* GLA.O, characterized in that, proceeding from the recombinant DNA molecule according to Claim 4,

10

- a) PCR primers are prepared,
- b) these PCR primers are used for accumulating DNA fragments of genomic DNA or cDNA from a corresponding microorganism,
- c) the accumulated fragments are isolated and characterized, and
- 15 d) where appropriate, employed in a process as described in the previous paragraph.

The described processes for isolating a gene cluster for the biosynthesis of acarbose and homologous pseudo-oligosaccharides from acarbose-producing microorganisms other than *Streptomyces glaucescens* GLA.O are characterized in that the microorganisms are selected from the group consisting of Actinomycetales, such as *Streptomyces*, *Actinoplanes*, *Ampullariella* and *Streptosporangium* strains, *Streptomyces hygroscopicus* var. *limoneus* and *Streptomyces glaucescens*, preferably from the group consisting of *Streptomyces glaucescens* GLA.O and *Actinoplanes* sp.

25

The invention furthermore relates to the use of *Streptomyces glaucescens* GLA.O for isolating acarbose.

30 The invention will now be explained in more detail with the aid of the examples, tables and figures, without being restricted thereto.

All the plasmid isolations were carried out using a Macherey and Nagel (Düren, Germany) isolation kit (Nucleobond[®]) in accordance with the manufacturer's instructions. Molecular biological procedures were carried out in accordance with standard protocols (Sambrock et al. (1989) Molecular cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, USA) or in accordance with the instructions of the respective

35

manufacturer. DNA and protein sequences were examined using Genetics Computer Group Software, Version 8 (programs: FastA, TFastA, BlastX, Motifs, GAP and CODONPREFERENCE) and the SwissProt (release 32), EMBL (release 46) and Prosite (release 12.2) databases. The molecular
5 biological manipulation of *Streptomyces glaucescens* and *Actinoplanes* (DNA isolation and DNA transformations) were carried out as described in Hopwood et al.: Genetic Manipulation of *Streptomyces*: A Laboratory Manual. The John Innes Foundation, Norwich, UK, 1985 and Motamedi and Hutchinson: Cloning and heterologous expression of a gene cluster for
10 the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of *Streptomyces glaucescens*. Proc. Natl. Acad. Sci. USA 84:4445-4449 (1987).

In general, hybridizations were performed using the "Non-radioactive DNA
15 labeling kit" from Boehringer Mannheim (Cat. No. 1175033). The DNA was visualized using the "Luminescent Detection Kit" from Boehringer Mannheim (Cat. No. 1363514). In all the examples given in this patent application, hybridization was carried out under stringent conditions: 68°C, 16 h. 5×SSC, 0.1% N-laurylsarcosine, 0.02% SDS, 1% Blocking Reagent
20 (Boehringer Mannheim). SSC denotes 0.15M NaCl/0.015M sodium citrate. The definition of "stringent conditions" which is given here applies to all aspects of the present invention which refer to "stringent conditions". In this connection, the manner of achieving this stringency, i.e. the cited hybridization conditions, is not intended to have a limiting effect since the
25 skilled person can select other conditions as well in order to achieve the same stringent conditions, e.g. by means of using other hybridization solutions in combination with other temperatures.

Example 1: Synthesis and sequences of the PCR primers and
30 amplification of the fragments from *S. glaucescens* GLA.O

The PCR was carried out under standard conditions using in each case 100 pmol of primer 1 and of primer 2 in 100 µl of reaction mixture

35	PCR buffer ¹	10 µl
	PCR primers	in each case 2.5 µl
	dNTPs	in each case 0.2 mM
	BSA (10 mg/ml)	1 µl

Template DNA 1 µg (1 µl)
Taq polymerase² (5 units/ml) 1.5 µl
H₂O to make up to 100 µl
1: Promega
5 2: Boehringer Mannheim

The samples are overlaid with 75 µl of mineral oil and the amplification is carried out using a Perkin Elmer TC1 DNA thermal cycler.

10 Parameters:

Cycles	Temperature	Duration
1	96°C	5 min
	74°C	5 min
30	95°C	1.5 min
	74°C	1.5 min
1	74°C	5 min

Table 1 lists the sequences of the degenerate primers which should be used for amplifying dTDP-glucose dehydratases from different streptomycetes.

15

Table 1: Primer sequences for amplifying dTDP-glucose 4,6-dehydratases

20 Primer 1: CSGGSGSSGCSGGSTTCATSGG (SEQ ID NO.: 1)

Primer 2: GGGWVCTGGYVSGGSCCGTAGTTG (SEQ ID NO.: 2)

In this table, S=G or C, W=A or T, V=A or G, and Y=T or C.

25

Example 2: DNA sequences of the PCR fragments isolated from *Streptomyces glaucescens* GLA.O

30 The sequencing was performed by the dideoxy chain termination method of Sanger et al. (PNAS USA, 74: 5463-5467 (1977)). The reactions were carried out using the Auto Read Sequencing Kit[®] from Pharmacia Biotech (Freiburg, Germany) in accordance with the manufacturer's instructions. An

ALF DNA Sequencer[®] from Pharmacia Biotech (Freiburg, Germany) was used for separation and detection.

The subsequent cloning of the PCR fragments (Sure Clone Kit[®],
5 Pharmacia Biotech, Frieburg) into the E. coli vector pUC 18, and the sequencing of the fragment, provided support for the supposition that the fragment encoded a dTDP-glucose 4,6-dehydratase. However, 2 different genes were isolated which both exhibit high degrees of homology with dTDP-glucose 4,6-dehydratase but are not identical. In that which follows,
10 the PCR fragments are designated acbD and HstrE.

The sequences of the isolated PCR fragments are shown in Table 2A and 2B and the homology comparison of the deduced amino acid sequences of HstrE and acbD is shown in Table 2C. The two proteins exhibit an identity
15 of only 65%.

Table 2A: DNA sequence of acbD (primer-binding sites are underlined, SEQ ID NO.: 3)

Primer 1

```
1  CCCGGGCGGG GCGGGGTTCA TGGGCTCCGC CTACGTCCGC CGGCTCCTGT
51 CGCCCGGGGC CCCCGGCGGC GTCGCGGTGA CCGTCCTCGA CAAACTCACC
101 TACGCCGGCA GCCTCGCCCG CCTGCACGCG GTGCGTGACC ATCCCGGCCT
151 CACCTTCGTC CAGGGCGACG TGTGCGACAC CGCGCTCGTC GACACGCTGG
201 CCGCGCGGCA CGACGACATC GTGCACTTCG CGGCCGAGTC GCACGTCGAC
251 CGCTCCATCA CCGACAGCGG TGCCTTCACC CGCACCAACG TGCTGGGCAC
301 CCAGGTCCTG CTCGACGCCG CGCTCCGCCA CGGTGTGCGC ACCCTCGTGC
351 ACGTCTCCAC CGACGAGGTG TACGGCTCCC TCCCGCACGG GGCCGCCGCG
401 GAGAGCGACC CCCTGCTCCC GACCTCGCCG TACGCGGCGT CGAAGGCGGC
451 CTCGGACCTC ATGGCGCTCG CCCACCACCG CACCCACGGC CTGGACGTCC
501 GGGTGACCCG CTGTTCTGAAC AACTACGGCC CGCACCAGTT CCCGGG
```

Primer 2

Table 2B: DNA sequence of HstrE^{*} (primer-binding sites are underlined, SEQ ID NO.: 4)

Primer 2

```

1  CCCCGGGTGC TGGTAGGGGC CGTAGTTGTT GGAGCAGCGG GTGATGCGCA
51  CGTCCAGGCC GTGGCTGACG TGCATGGCCA GCGCGAGCAG GTCGCCCCGAC
101 GCCTTGGAGG TGGCATAGGG GCTGTTGGGG CGCAGCGGCT CGTCCTCCGT
151 CCACGACCCC GTCTCCAGCG AGCCGTAGAC CTCGTCGGTG GACACCTGCA
201 CGAAGGGGGC CACGCCGTGC CGCAGGGCCG CGTCGAGGAG TGTCTGCGTG
251 CCGCCGGCGT TGCTCCGCAC GAACGCGGGC GCATCGAGCA GCGAGCGGTC
301 CACGTGCGAC TCGGCGGCCA GGTGCACGAC CTGGTCCTGG CCGGCCATGA
351 CCCGGTCGAC CAGGTCCGCG TCGCAGATGT CGCCGTGGAC GAAGCGCAGC
401 CGGGGGTGGT CGCGGACCGG GTCGAGGTTG GCGAGGTTGC CGGCGTAGCT
451 CAGGGCGTCG AGCACGGTGA CGACGGCGTC GGGCGGCCCC TCCGGACCGA
501 GGAGGGTGCG GACGTAGTGC GAGCCCATGA ACCCGCCCCC C

```

Primer 1

5

Table 2C: Homology comparison of the deduced amino acid sequences of the PCR products HstrE^{*} and acbD^{*} (program: GAP)

Quality:	196.3	Length:	182
Ratio:	1.091	Gaps:	0
Percent similarity:	77.654	Percent identity:	65.363

10 PCRstrE.Pep × PCRaCbD.Pep

```

1  ..AAGFMGSHYVRTLLGPDGPPDAVVTVLDALSYAGNLANLDPVRDHPRL 48
   :|||:| | | | | :| :| :| :| :| :| :| :| :| :| :| :| :|
1  PGGAGFIGSAYVRRLLSPGAPGGVAVTVLDKLTAGSLARLHAVRDHPGL 50
49 RFVHGDICDADLVDRVMAGQDQVVHLAAESHVDRSLLDAAAFVRTNAGGT 98
   ||:|:|:| |..| | | : | :| :| :| :| :| :| :| :| :| :|
51 TFVQGDVCDTALVDTLAARHDDIVHFAAESHVDRSITDSGAFTRTNVLGT 100
99 OTLLDAALRHGVAPFVQVSTDEVYGSLETGSWTEDEPLRPNSPYATSKAS 148
   |. | | | | | | | | | | | | | | | | | | | | | | | | | |
101 QVLLDAALRHGVRTLHVSTDEVYGS LPHGAAAESDPIPTSPYAASKAA 150
149 GDLLALAMHVSHGLDVRITRCSNNYGPYQHPG 180
   :|:|:| | | | | | | | | | | | | | | | | | | | |
151 SDLMAHAHRTHGLDVRVTRCSNNYGPHQFP. 181

```

in each case, upper row: SEQ ID NO.: 5

in each case, lower row: SEQ ID NO.: 6

Example 3: Southern analysis using chromosomal DNA from *Streptomyces glaucescens* GLA.O and the isolated and labeled PCR fragments

5 The cells were grown in R2YENG medium and harvested for the DNA isolation after 30 h. The chromosomal DNA was isolated from *S. glaucescens* GLA.O as described in Hopwood et al. (1985) Genetic manipulations of *Streptomyces*: a laboratory manual. The John Innes Foundation, Norwich UK).

10

A Southern blot analysis was carried out using the *S. glaucescens* GLA.O producer strain chromosomal DNA, which was digested with *Pst*I, *Bgl*II and *Bam*HI, using the labeled probes consisting of the *acbD* and *HstrE* PCR fragments. The two PCR fragments were labeled with digoxigenin in accordance with the manufacturer's (Boehringer Mannheim; Mannheim) instructions, and a digest of the *Streptomyces glaucescens* GLA.O producer strain chromosomal DNA was separated on an agarose gel. The DNA was transferred by capillary transfer to nylon membranes and DNA regions which were homologous with the labeled probes were subsequently visualized following hybridization.

The two genes label different DNA regions (Fig. 1 and Fig. 2), with the fragments which were labeled by *HstrE* having to be gene fragments from *Streptomyces glaucescens* GLA.O hydroxystreptomycin biosynthesis. While the DNA sequence is not published, the high degree of homology of the protein sequence deduced from *HstrE* with *StrE* (Pissowotzki et al. (1991) Mol. Gen. Genet. 231: 113-123) from *Streptomyces griseus* N2-3-11 streptomycin biosynthesis (82% identity) and the concordance of the *HstrE* -labeled DNA fragments (Fig. 1) with the published restriction map of the *Streptomyces glaucescens* GLA.O hydroxystreptomycin gene cluster (Retzlaff et al. (1993) Industrial Microorganisms. Basic and applied molecular genetics ASM, Washington DC, USA) permits this conclusion. The fragments which were labeled by the *acbD* probe (Fig. 2) belong to a DNA region which has not previously been investigated. This region encodes the enzymes for biosynthesizing the *Streptomyces glaucescens* GLA.O pseudo-oligosaccharides.

Example 4: Cloning the 6.8 kb PstI fragment

Inter alia, the *acbD* PCR fragment labels a 6.8 kb PstI DNA fragment (Fig. 2). This DNA fragment was isolated as follows. The region of the gel
5 was excised with a razor blade and the DNA was isolated from the gel using an isolation kit from Pharmacia Biotech and cloned into plasmid pUC19 which had been cut with the restriction enzyme PstI (plasmid *pacb1*); this latter plasmid was then transformed into the *E. coli* strain DH5 α . The individual clones were subcultured from these plates and a
10 plasmid DNA isolation was carried out using these clones. A PCR amplification using the above-described primers 1 and 2 (Tab. 1) was carried out using the DNA from these clones (250). In this manner, the appropriate *E. coli* clone containing the 6.8 kb PstI fragment was isolated.

15 Example 5: Sequencing the isolated 6.8 kb PstI DNA fragment

The DNA was digested with various restriction enzymes and individual DNA fragments were cloned into pUC19. The DNA sequence of the entire fragment, which is shown in Tab. 4 (SEQ ID NO.: 7), was then determined.
20 The DNA sequence of the 6.8 kb PstI fragment was only partially confirmed by supplementary sequencing of the opposing strand. Several open reading frames, encoding various proteins, were found (programs: CODONPREFERENCE and BlastX). A total of 6 coding regions was found, i.e. a gene having a high degree of homology with ATP-binding protein, *acbA*, an aminotransferase *acbB*, a dTDP-glucose synthase *acbC*, a
25 dTDP-glucose dehydratase *acbD*, a regulatory gene having homologies with the LacI protein family *acbE*, and a protein having similarities to sugar-binding proteins *acbF*. The sequences of the *acbA* and *acbF* genes were only determined in part. The homologies with other proteins from the
30 databases, and the properties of the putative proteins, are summarized in Tab. 3. Fig. 3 shows, in summary form, a restriction map of the fragment, containing the most important restriction cleavage sites mentioned in the text, and the arrangement of the identified open reading frames.

Table 3: Analysis of the identified open reading frames on the 6.8 kb PstI fragment from *Streptomyces glaucescens* GLA.O

ORF	Amino acid	MW	FastA [§]	%Identity	Accession number [§]
acbA	239	.	MalK E coli	29%	P02914
acbB	429	45618	DgdA, Burkholderia cepacia	32%	P16932
acbC	355	37552	StrD, Streptomyces griseus	60%	P08075
acbD	325	35341	StrE, Streptomyces griseus	62%	P29782
acbE	345	36549	DegA, Bacillus subtilis	31%	P37947
acbF	396	.	MalE, E. coli	22%	P02928

5 * incomplete open reading frame; [§] Swiss-Prot database (release 32)

Example 6: Deletion of genes acbBCD for pseudo-oligosaccharide biosynthesis from the *Streptomyces glaucescens* GLA.O chromosome

10

Evidence that the identified DNA fragment encoded pseudo-oligosaccharide biosynthesis genes was produced as follows. A 3.4 kb gene region (EcoRI/SstI fragment b, Fig. 3) was replaced with the erythromycin resistance gene (1.6 kb) and cloned, together with flanking DNA regions from the 6.8 kb PstI fragment (pacb1) into the temperature-sensitive plasmid pGM160. The plasmid was constructed as described in the following: the 2.2 kb EcoRI/HindIII fragment (c, Fig. 3) from plasmid pacb1 was cloned into pGEM7zf (Promega, Madison, WI, USA; plasmid pacb2), and the 1 kb SstI fragment from pacb1 (a, Fig. 3) was cloned into pUC19 (plasmid pacb3). A ligation was then carried out using the following fragments. The plasmid pGM160 (Muth et al. (1989) Mol. Gen. Genet. 219:341-348) was cut with BamH/HindIII, the plasmid pacb2 was cut with XbaI/BamHI (c, Fig. 3), the plasmid pacb3 was cut with EcoRI/HindIII (a, Fig 3), and the plasmid pIJ4026 (Bibb et al. (1985) Gene 38:215-226) was cut with EcoRI/XbaI in order to isolate the 1.6 kb ermE resistance gene.

15

20

25

The fragments were ligated in a mixture and transformed into *E. coli* DH5 α and selected on ampicillin. The resulting plasmid, i.e. *pacb4*, was isolated from *E. coli* DH5 α , tested for its correctness by means of restriction digestion and then transferred by protoplast transformation into

5 *S. glaucescens* GLA.O. The transformants were selected with thiostrepton at 27°C in R2YENG agar. The transformants were subsequently incubated at the non-permissive temperature of 39°C and integration of the plasmid into the genome by way of homologous recombination thereby instituted (selection with thiostrepton (25 μ g/ml) and erythromycin (50 μ g/ml)). Under

10 these conditions, the only clones which can grow are those in which the plasmid has become integrated into the genome. The corresponding clones were isolated, caused to sporulate (medium 1, see below) and plated out on erythromycin-containing agar (medium 1). Individual clones were isolated once again from this plate and streaked out on both

15 thiostrepton-containing medium and erythromycin-containing medium. The clones which were erythromycin-resistant but no longer thiostrepton-resistant were analyzed. In these clones, the *acbBCD* genes had been replaced with *ermE*. Several clones were examined and the strain *S. glaucescens* GLA.O Δ *acb* was finally selected as the reference strain

20 (erythromycin-resistant, thiostrepton-sensitive) for further investigation.

Medium 1

	Yeast extract	4 g/L
25	Malt extract	10 g/L
	Glucose	4 g/L
	Agar	15 g/L
	pH	7.2

30 A further experiment examined whether the corresponding strain still produced acarbose. Some clones were grown and investigated for formation of the α -amylase inhibitor in a bioassay; however, no activity was found. The mutants were subsequently further characterized by means of Southern hybridization. Integration of the *ermE* gene had taken place at

35 the predicted site. Fig. 4 shows a Southern hybridization which was carried out with the wild type and with the *Streptomyces glaucescens* GLA.O Δ *acb* deletion mutant. The *Sst*I fragment from *pacb3* was used as the probe. The chromosomal DNA was isolated from the wild type and mutant and

digested with the enzymes PstI and PstI/HindIII. The fragment pattern obtained for the deletion mutant corresponds to the predicted recombination event. The wild type exhibits the unchanged 6.8 kb PstI fragment, whereas the mutant exhibits a fragment which has been
5 truncated by 1.8 kb (compare lanes 1 and 3, Fig. 4). Integration of the ermE resistance gene additionally introduced an internal HindIII cleavage site into the PstI fragment (compare lanes 2 and 4, Fig. 4).

Example 7: Inhibition of α -amylase by acarbose

10

Using an enzymic test for detecting starch (TC-Starch, Boehringer-Mannheim, Cat. No. 297748), it was possible to demonstrate that the compound isolated from *Streptomyces glaucescens* GLA.O inhibits α -amylase. Test principle: starch is cleaved into D-glucose by
15 amyloglucosidase. The glucose is then converted with hexokinase into glucose-6-phosphate and the latter is converted with glucose-6-phosphate dehydrogenase into D-gluconate-6-phosphate. This reaction produces NADPH, whose formation can be determined photometrically. Acarbose inhibits the α -amylase and thereby prevents the formation of D-glucose
20 and ultimately the formation of NADPH as well.

Example 8: Medium for growing *S. glaucescens* GLA.O and producing acarbose

25 The fermentation was carried out, at 27°C on an orbital shaker at 120 rpm, in 500 ml Erlenmeyer flasks which were fitted with side baffles and which contained 100 ml of medium 2. The fermentation was terminated after 2 or 3 days. The pseudo-oligosaccharides were detected in a plate diffusion test as described in Example 9. No α -amylase inhibitors were produced
30 when medium 3 was used. This means that the production of the pseudo-oligosaccharides is inhibited by glucose. Other sugars, such as maltose and sucrose, or complex sugar sources (malt extract) can also come into consideration for producing pseudo-oligosaccharides using *S. glaucescens* GLA.O.

35

Medium 2:

Soybean flour 20 g/L

Starch	20 g/L
pH	7.2

Medium 3:

5

Soybean flour	20 g/L
Glucose	20 g/L
pH	7.2

10 Example 9: Biotest using *Mucor miehei*

A suspension of spores of the strain *Mucor miehei* was poured into agar (medium 5) (10^5 spores/ml), and 10 ml of this mixture were in each case poured into Petri dishes. Paper test disks (6 mm diameter) were loaded with 10 μ l of acarbose [lacuna] (1 mg/ml) or with a sample from an *S. glaucescens* culture and laid on the test plates. The plates were then incubated at 37°C. Inhibition halos appeared on the starch-containing medium 5. A plate which was prepared with glucose (medium 4) instead of starch was used as a control. On this medium, no inhibition halo formed around the filter disks loaded with active compound.

Media 4 and 5:

	KH ₂ PO ₄ × 3 H ₂ O	0.5 g
25	MgSO ₄ × 7 H ₂ O	0.2 g
	NaCl	0.1 g
	Ammonium sulfate	5 g
	Yeast nitrogen base	1.7 g
	Glucose (4) or starch (5)	5 g
30	Agar	15 g

Example 10: Transformation of *S. glaucescens* GLA.O

Protoplasts of the *Streptomyces glaucescens* strain were isolated as described in Motamedi and Hutchinson ((1987) PNAS USA 84: 4445-4449), and the isolated plasmid DNA was transferred into the cells by means of PEG transformation as explained in Hopwood et al. ((1985) Genetic manipulations of *Streptomyces*: a laboratory manual. The John

Innes Foundation, Norwich UK). The protoplasts were regenerated on R2YENG medium at 30°C (Motamedi and Hutchinson (1987) PNAS USA 84: 4445-4449). After 18 h, the agar plates were overlaid with a thiostrepton-containing solution and incubated at 30°C (final concentration of thiostrepton: 20 µg/ml).

Example 11: Isolation of the pseudo-oligosaccharides from *Streptomyces glaucescens* GLA.O, HPLC analysis and mass spectroscopy

10 Isolation

The culture broth was separated from the mycelium by filtration. The culture filtrate which has been obtained in this way is then loaded onto an XAD16 column, after which the column is washed with water and the active components are eluted with 30% methanol. The eluate was concentrated down to the aqueous phase and the latter was extracted with ethyl acetate in order to remove lipophilic impurities. The aqueous phase was then concentrated and the active components were further purified in 5% methanol using a biogel P2 column. The individual fractions are collected in a fraction collector. The individual fractions were analyzed by means of the *Mucor miehei* biotest. Active eluates were rechromatographed, for further purification, in 5% methanol on biogel P2. The material which was isolated in this way was investigated by HPLC and MS.

25 HPLC

Column: Nucleosil[®] 100 C-18

Eluent 0.1% phosphoric acid = A/acetonitrile = B

Gradient: from 0 to 100% B in 15 min

30 Detection: 215 nm

Flow 2 ml/min

Injection volume: 10-20 µl

Using HPLC, it was not possible to distinguish the pseudo-oligosaccharide preparation from *S. glaucescens* GLA.O from authentic acarbose. Both the retention time and the UV absorption spectrum of the two components were identical in this eluent system. The pseudo-oligosaccharide mixture was not fractionated under these conditions.

Mass spectroscopic analysis (MS)

The molecular weights and the fragmentation pattern of authentic acarbose and the pseudo-oligosaccharides isolated from *Streptomyces glaucescens* GLA.O were determined by means of electrospray MS. Analysis of the acarbose which is commercially obtainable from Bayer (Glucobay) gave a mass peak at 645.5 (acarbose). The purified samples from *S. glaucescens* GLA.O contain a mixture of different pseudo-oligosaccharides whose sugar side chains are of different lengths: 969 (acarbose + 2 glucose units), 807 (acarbose + 1 glucose unit), 645 (corresponds to authentic acarbose). When acarbose and the compound which is isolated from *S. glaucescens* GLA.O and which has a molecular weight of 645 are fragmented, the same molecular fragments are formed, i.e.: 145 (4-amino-4,6-deoxyglucose), 303 (Acarviosin) and 465 (303 together with one glucose unit).

Actinoplanes sp. SE50 also produces a mixture of acarbose molecules having sugar side chains of different length (Truscheit (1984) VIIIth International Symposium on Medicinal Chemistry, Proc. Vol 1. Swedish Academy of Pharmaceutical Sciences, Stockholm, Sweden). The length of the sugar side chains can be influenced by the choice of the fermentation parameters and of the substrate in the nutrient solution.

Example 12: Southern hybridization using *Actinoplanes* sp. SE50/110 (ATCC31044)

The chromosomal DNA was isolated from the strain *Actinoplanes* sp. SE50/100 and digested with restriction enzymes (PstI and BamHI). A Southern hybridization was then carried out using a probe which encompasses the coding region of the dTDP-glucose 4,6-dehydratase *acbD* from *Streptomyces glaucescens* GLA.O (fragment d, Fig. 3). The probe hybridizes with distinct bands from *Actinoplanes* sp. SE50/110 (Fig. 5, lanes 1 and 2). This provides the possibility of isolating the corresponding fragments from *Actinoplanes* sp. SE50/100 and other strain lines. Whether these DNA regions are in fact involved in the biosynthesis of acarbose remains to be demonstrated in subsequent investigations. Alternatively, the PCR primers 1 and 2 (Tab. 1) could also be used for amplifying the dTDP-glucose 4,6-dehydratase from *Actinoplanes* sp.

Legends:

- 5 Fig. 1: Southern hybridization using *S. glaucescens* GLA.O. Lane 1: PstI, lane 2: BamHI, lane 3: BglII. The labeled PCR fragment HstrE was used as the probe. Labeling of DNA fragments which are involved in the biosynthesis of hydroxystreptomycin.
- 10 Fig. 2: Southern hybridization using *S. glaucescens* GLA.O. Lane 1: PstI, lane 2: BamHI, lane 3: BglII. The labeled PCR fragment acbD was used as the probe. Labeling of DNA fragments which are involved in the biosynthesis of the pseudo-oligosaccharides.
- 15 Fig. 3: Restriction map of the 6.8 kb PstI fragment from *Streptomyces glaucescens* GLA.O. Open reading frames and the direction in which each is transcribed are indicated by arrows. The fragments a, b, c and d identify DNA regions which are explained in more detail in the text.
- 20
- 25 Fig. 4: Southern hybridization using *Streptomyces glaucescens* Δ acb: lane 1: PstI, lane 2: PstI/HindIII, and *Streptomyces glaucescens* GLA.O lane 3: PstI, lane 4: PstI/HindIII. The labeled 1.0 kb SstI fragment a (Fig. 3) was used as the probe.
- 30 Fig. 5: Southern hybridization using *Actinoplanes* sp. SE50/100: lane 1: PstI, lane 2: BamHI and *Streptomyces glaucescens* GLA.O lane 3: PstI. The labeled 1.0 kb SmaI/EcoRI fragment d (dTDP-glucose 4,6-hydratase, Fig. 3) was used as the probe. The arrows indicate the labeled DNA fragments (BamHI: 2.1 and 0.7 kb, PstI: ~11-12 kb)
- 35 Tab. 4: DNA sequence of the 6.8 kb PstI fragment from *Streptomyces glaucescens* GLA.O (SEQ ID NO.: 7). The deduced amino acid sequences (SEQ ID NO.: 8-13) of the identified open reading frames are given under the DNA

sequences. Start and stop codons and potential ribosome binding sites are underlined.

acbA: SEQ ID NO.: 8

acbB: SEQ ID NO.: 9

5 acbC: SEQ ID NO.: 10

acbD: SEQ ID NO.: 11

acbE: SEQ ID NO.: 12

acbF: SEQ ID NO.: 13

Table 4: (SEQ ID NO.: 7, 8, 9, 10, 11, 12, 13)

P
S
t
I

CTGCAGGGTTCCCTGGTGCACGACCGCCCTGGTGCAGCAGCAGGGCGCTGTCCGAGAT
-----+-----+-----+-----+-----+ 60
GACGTCCCAAGGGACCACTGTCTGGGCGGGACCAGCTGCTGGTCCCGCGACAGCGTCTA
Q L T G Q H V V R G Q D V V L A S D C I -
CGCGCGGATGTCCGCCATGTCGTGGCTGGTGAGCACCACGGTGGTGCCAGTTCCCGGTG
-----+-----+-----+-----+-----+ 120
GCGCCGCTACAGCCGCTACAGCACCGACCACTCGTGGTGCCACCACGGGTCAAGGGCCAC
A A I D A I D H S T L V V T T G L E R H -
GGCGCGGTTGACCAGCCGGCGCACC CGCTCCTTCAGCACCATGTGAGGCCGATCGTGGG
-----+-----+-----+-----+-----+ 180
CCGCGCCAACCTGGTCCGCGCGTGGCGCAGGAAGTCTGTGGTACAGCTCCGGCTAGCACCC
A R N V L R R V A D K L V M D L G I T P -
CTCGTCCCAGAACAGCACGGCCGGGTCTGTGCAGCAGGCTCGCCGCGATCTCGGCCGCGCAT
-----+-----+-----+-----+-----+ 240
GAGCAGGCTCTTGTCTGTGCGCGGCCAGCACGTCTGTCCGAGCGGCGCTAGAGCCGCGCGTA
E D W F L V A P D H L L S A A I E A R M -
S
P
h
I
CGCTGTCCGAGGCTGAGCTGCCGACGGGGGTGGACCCAGCGCGTCTGATGTGAGGAG
-----+-----+-----+-----+-----+ 300
CGCGACAGGCTCCGACTCGACGGCGTCCCCCACCTGGGGTCCGCGCAGCTACAGCTCCTC
R Q G L S L Q R V P T S G L A D I D L L -
GTCCCGGAACAGGGCGAGGTTGCCCGGTAGACCGGTCCGGGGATGTCGTAGATGCCGGCG
-----+-----+-----+-----+-----+ 360
CAGGGCCTTGTCCCGCTCCAACGCGGCCATCTGGCCAGGCCCTACAGCATCTACGCCGC
D R F L A L N R R Y V P G P I D Y I R R -
K
P
n
I
CAGGATGCCGAAGGAGTCCGGTACCGACAGGTCCCACCAGAGCTGGCTGCGCTGGCCGAA
-----+-----+-----+-----+-----+ 420
GTCTACGCCTTCCTCAGCCCATGGCTGTCCAGGGTGGTCTCGACCGACGCGACCGGCTT
L I R F S D P V S L D W W L Q S R Q G F -
CAGCAGCGCCGATCGTGGGGCGTTGCGCTGCCGGTGCCGGTAGGGCTCCAGCCCGGCGAC
-----+-----+-----+-----+-----+ 480
CTGCTGCGGCTAGCACGCCCCGAACGCGACGGCCACGGCCATCCCGAGGTCCGGCCGCTG
V V G I T R A N R Q R H R Y P E L G A V -
CGTGACGCGGCCGGAGGTGGGGGTCTATGATGCCGGTCAGCATCTTGATCGTGGTGCAGTT
-----+-----+-----+-----+-----+ 540
GCACGTCCGCCGGCTCCACCCCCAGTACTACGGCCAGTCTGTAAGACTAGCACCAGCTGAA
T C R G S T P T M I G T L M K I T T S K -
GCCCCCTCCGTTGGCGCCGATGTAGGCGGTCTTCGTGCCGGCCGGTATCTCGAAGGAGAC
-----+-----+-----+-----+-----+ 600
CGGCCGAGGCAACCGCGGCTACATCCGCCAGAAGCACGGCCCGCCATAGAGCTTCCTCTG
G A G N A G I Y A T K T G A P I E F S V -

K
P
n
I

GTGTCGACCGGCGCGCACGACGCGGTACCGGCGGGTCAGGAGGGTGGAGAGGCTGCCGAG
-----+-----+-----+-----+-----+ 660
CAGCAGCTGCCGCGCGTCTGCGCCATGGCGCCCACTCCTCCCACTCTCCGACGGGCTC
D D V A R V V R Y R R T L L T S L S G L -

CAGGCGGGGCTCGCGTTCGGCCAGCGGAACTCCTTGACGAGGTGTTCCGCCACCATCAC
-----+-----+-----+-----+-----+ 720
GTCCGGCCCCGAGCGCAAGCGGTGCGGCCTTGAGGAACTGCTCCACAAGCGGTGCTAGTG
*
L G P E R E A L R P E K V L H E A V I V -
acbA
GCGATCACCCGCTCGACGGCCGTCTCCAGCAGGGCGCAGGCCCTCGTCGAGCAGCGCCTCG
-----+-----+-----+-----+-----+ 780
CGCTAGTGGGCGAGCTGCCGGCAGAGGTCTGTCGCGTCCGGGAGCAGCTCGTCGCGGAGC
A I V R E V A T E L L R L G E D L L A E -

TCGAGGGTGAACGGCGGTGCCAGCCGCGAGGATGTGGCCGCCAGGGAGGTGCCGAGCCCC
-----+-----+-----+-----+-----+ 840
AGCTCCCACTTGCCGCCACGGTCCGGCTCTACACCGGCGGGTCCCTCCACGCGTCGGGG
D L T F P P A L R L I H G G L S T R L G -

S
m
a
I

AGGTCGAGGGCGGTGGTGTAGACGGCCCGGGCGGTCTCGGGGGCGGGTGCCCGGCCGACG
-----+-----+-----+-----+-----+ 900
TCCAGCTCCCGCCACCACATCTGCCGGGCGCCGAGAGCCCCCGCCACGGGCGGGTGC
L D L A T T Y V A R A T E P A P A R G V -

GCGTCGGTGACGAACTCCAGGCCCCACAGCAGTCCGAGGCGCGTACCTGGCCGAGCTGG
-----+-----+-----+-----+-----+ 960
CGCAGCCACTGCTTGAGGTCCGGGGTGTCTGAGGCTCCGGCGCATGGACCGGCTCGACC
A D T V F E L G W L L G L G R V Q G L Q -

S
s
t
I

GGGAAGCGGGACTCCAGGGCGCGCAGCCGCTCCTGGATGAGCTCGCCGAGGACGCGCACG
-----+-----+-----+-----+-----+ 1020
CCCTTCGCCCTGAGGTCCCGCGCGTCCGGCAGGACCTACTCGAGCGGCTCCTGCGCGTGC
P F R S E L A R L R E Q I L E G L V R V -

CGGTCGATCAGCCGGTCCGCGCTCGACGACCTCCAGCGTGGCGCGGGCGGGCGGATCCCC
-----+-----+-----+-----+-----+ 1080
GCCAGCTAGTCGGCCAGCGCGAGCTGCTGGAGGTGCGACCGCGCCCGCCGCGCTAGGGG
R D I L R D R E V V E L T A R A A A I G -

S
m
a
I

AGTGGGTTGCTCGCGTACGTCGAGGCGTACGCCCCGGGGTGGCCGCCTCCGGCCTGCGCA
-----+-----+-----+-----+-----+ 1140
TCACCCAACGAGCGCATGCAGCTCCGCATGCGGGGCCCCACCGGCGGAGGCCGACGCGT
L P N S A Y T S A Y A G P H G G G A Q A -

GCTTCCGCGCGTCCGGCCAGCACGGCGAAGGGGAATCCGCTCGCGGTGCCCTTGGACAGC
-----+-----+-----+-----+-----+ 1200
CGAAGGCGCGCAGGCCGGTCTGTGCCGCTTCCCCTTAGGCGAGCGCCACGGGAACCTGTCTG
A E A R G A L V A F P F G S A T G K S L -
ATCGCCAGGTCCGGCTCGATGCCGAACAGTTCTGCTGGCGAGGAAGGCGCCGGTGC GCCCG
-----+-----+-----+-----+-----+ 1260
TAGCGGTCCAGGCCGAGCTACGGCTTGTCAAGCGACCGCTCCTTCCGCGGCCACGCGGGC
M A L D P E I G F L E S A L F A G T R G -
CCGCGCGGTGAGGACCTCGTCCGGCGACGAGCAGCACGCCGCGCTCCCGGCCAGGCGCCGGCG
-----+-----+-----+-----+-----+ 1320
GGCGGCCACTCCTGGAGCAGCCGCTGCTCGTCTGCGGCGGCAGGGCCCTCCGCGGCCCGC
G G T L V E D A V L L V G G D R C A G A -
ATCCGCTCCCACTAGCCGGGGGGCGGCACGATGACGCTGCCGCGCCGAGGACGGGTTCTG
-----+-----+-----+-----+-----+ 1380
TAGGCGAGGGTCTCGGCCCGCGCGCTGCTACTGCGGACGGCGCGGCTCCTGCCCAAGC
I R E W Y G P P P V I V G A A G L V P E -
AAGACCAGGGCCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCGCACGAGGGTGC GCGCAC
-----+-----+-----+-----+-----+ 1440
TTCTGGTCCCGGCTCTGCAACCCGAAGAGGCGCTACACGGCCGCGTGTCTCCAGCGCGTG
F V L A S V N P K E A I H R R V L T A C -
CGCACGTGCGACGAGGGGTACTCCAGGCCAGGGGACAGCGGTAGCCAGTAGGGGCTGTA
-----+-----+-----+-----+-----+ 1500
GCGTGACGCGTGCTCCCATGAGGTCCGGGTCCCCTGTGCGCATCGGTCTCCCGACAT
R V D C S P Y E L G L P C R Y G T P A T -
GCCAGCACGCTGTTGCCGCTGAAGGCCTGGTGGCCGATGTCCAGTGGACCAGCATCCGG
-----+-----+-----+-----+-----+ 1560
CGGTCTGTGCGACAACGGCGACTTCCGGACCACCGGCTACAGGGTCACTGGTCTAGGCC
A L V S N G S F A Q H G I D W H V L M R -
GCGCCCATGGTCTTGCCGTGGAAGCCGTGGCGCAGGGCGCAGATCCGGTTGCGGCCCGGC
-----+-----+-----+-----+-----+ 1620
CGCGGGTACCAGAACGGCACCTTCCGGCACCGCGTCCCGCGTCTAGGCCAACGCCGGGCCG
A G M T K G H F G H R L A C I R N R G P -
GCGGCGGTGCGCTGGACGACCCGCGAGGGCGGCCTCGACCACCTCCGCGCCGGTGGAGAAG
-----+-----+-----+-----+-----+ 1680
CGCCGCCAGCGGACCTGCTGGGCGTCCCGCCGGAGCTGGTGGAGGCGCGGCCACCTCTTC
A A T A Q V V R L A A E V V E A G T S F -
AAGGCGTAGGTGTGAGCTGTTCCGGCAGCAGCCTGGCGAGCAGTTCCAGCAGGCCGGCG
-----+-----+-----+-----+-----+ 1740
TTCCGCATCCACAGCTCGACAAGCCCGTCTGTCGGACCGCTCGTCAAGGTCTGTCGGCCCGC
F A Y T D L Q E P L L R A L L E L L G A -
CGGTCCGGCGTGGCGCTGTCTGTGACGTTCCACAGGCGGGCGGCCTGGGTGGTGAGTGCC
-----+-----+-----+-----+-----+ 1800
GCCAGGCCGACCGCGACAGCACCTGCAAGGTGTCCGCCGCCCGGACCCACCTCACGG
R D P T A S D H V N W L R R A Q T T L A -
TCGACGACCTCCGGGTGCCCGTGGCCAGTGA CTGGGTGAGGGTCCCGGCCGCGAAGTCTG
-----+-----+-----+-----+-----+ 1860
AGCTGCTGGAGGGCCACGGGCACCGGGTCACTGACCCACTCCAGGGCCGGCGCTTCAGC
E V V E P H G H G L S Q T L T G A A F D -

AGGTACTGGTTGCCGTCCAGGTCGGTCAGAACGGGACCGGTCCTCGGCGAAGACCCGG
-----+-----+-----+-----+-----+ 1920
TCCATGACCAACGGCAGGTCAGCCAGTCTTGCCCTGGCGCAGGGAGCCGCTTCTGGGCC
L Y Q N G D L D T L V P G R G E A F V R -
CGTCCGTGGACGGCTTCCTCGGAGGGCGCCCGGCCAGGTGGCGGGCCTCCCGTGCCAGG
-----+-----+-----+-----+-----+ 1980
GCAGGCACCTGCCGAAGGAGCCTCCGCGGGCCGCGGTCCACCGCCCGGAGGGCACGGTCC
R G H V A E E S A G P A L H R A E R A L -
TGCTGTGTCTGCCGTAAGCCTGTCATCGCTGCCTCTGCTCGTCCGACCGGCTGACGCGAT
-----+-----+-----+-----+-----+ 2040
ACGACACAGACGGCATTTCGGACAGTAGCGAGACGAGCAGCCTGGCCGACTGCGCTA
H Q T Q R L G T M
-----+-----+-----+-----+-----+
CGCCGGCGAACTGCGTTGTGGCGCACCAAGGTTGGGGCGGCTCGGCGCTGAGTCAAACAC
-----+-----+-----+-----+-----+ 2100
GCGGCCGCTTGACGCAACACCGGTGGTGCCAAACCCCGCCGAGCCGCGACTCAGTTTGTG
TGGAACACACACCGCTGCAAGAGTTTGC CGGTTGTTTCAGAAAGTTGTTGCGAGCGGCCC
-----+-----+-----+-----+-----+ 2160
AACTTGTGTGTGGCGACGTTCTCAAACGCCCAACAAAGTCTTTCAACAACGCTCGCGGG
CGGCACTCTGGTTGAGTCGACGTGCTTACGGCGCCACCAACGCTCAGTTTCGAGGAGGGA
-----+-----+-----+-----+-----+ 2220
GCCGTGAGACCAACTCAGCTGCACGAATGCCCGCGTGGTGCGGAGTGCAAGCTCCTCCCT
CCTGTGAGAACAAGCCCGCAGACCGACCCGCTCCCGCGGAGGCGGAGGTGAAGGCCCTGG
-----+-----+-----+-----+-----+ 2280
GGACACTCTTGTTCGGCGCTCTGGCTGGGCGAGGGCGCCTCCGGTCCACTTCCGGGACC
V K A L V -
-----+-----+-----+-----+-----+
acbc
-----+-----+-----+-----+-----+
P
V
u
I
I
TCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTACCCACACCGCCGCCAAGCAGC
-----+-----+-----+-----+-----+ 2340
AGGACCGTCCACCTTGGCCGTCGTCTGACTCCGGCAAGTGGGTGTGGCGGCGGTTCTGTCG
L A G G T G S R L R P F T H T A A K Q L -
TGCTCCCATCGCCAACAAGCCCGTGCTCTTCTACGCGCTGGAGTCCCTCGCCGCGGCGG
-----+-----+-----+-----+-----+ 2400
ACGAGGGGTAGCGTTGTTCCGGCAGGAGAAGATGCGCGACCTCAGGGAGCGGCGCGCC
L P I A N K P V L F Y A L E S L A A A G -
GTGTCCGGGAGGCCGGCGTCTGCTGGGCGGTACGGCCGGGAGATCCGCGAACTCACCG
-----+-----+-----+-----+-----+ 2460
CACAGGCCCTCCGGCCGAGCAGCACCCGCGCATGCGGCCCTCTAGGCGCTTGAAGTGGC
V R E A G V V V G A Y G R E I R E L T G -
GCGACGGCACCGGTTCCGGTTACGCATCACCTACCTCCACCAGCCCCGCGCGCTCGGTG
-----+-----+-----+-----+-----+ 2520
CGCTGCCGTGGCGCAAGCCCAATGCGTAGTGGATGGAGGTGGTCCGGGGCGGGCGAGCCAG
D G T A F G L R I T Y L H Q P R P L G L -
TCGCGCACGCGGTGCGCATCGCCCGCGGCTTCCTGGGCGACGACGACTTCCTGCTGTACC
-----+-----+-----+-----+-----+ 2580
AGCGCGTGCGCCACGCGTAGCGGGCGCCGAAGGACCCGCTGCTGCTGAAGGACGACATGG
A H A V R I A R G F L G D D D F L L Y L -

TGGGGGACAACTACCTGCCCCAGGGCGTCACCGACTTCGCCCCGCAATCGGCCGCCGATC 2640
-----+-----+-----+-----+-----+
ACCCCTGTGATGGACGGGGTCCCGCAGTGGCTGAAGCGGGCGGTAGCCGGCGGCTAG
G D N Y L P Q G V T D F A R Q S A A D P -
COGCGGCGGCGCGGCTGCTGCTCACCOCGGTCGCGGACCGCTCGGCCTTCGGCGTCCGGG 2700
-----+-----+-----+-----+-----+
CGCGCGGCGGCGGCGGACGACGAGTGGGGCCAGCGCCTGGGCAGGCGGAAGCCGCGCGCC
A A A R L L L T P V A D P S A F G V A E -
AGGTGACGCGGACGGGAACGTGCTGCGCTTGGAGGAGAAACCGACGTCCCGCGCAGCT 2760
-----+-----+-----+-----+-----+
TCCAGCTGCGCCTGCCCTTGACGACGCGAACCTCCTCTTTGGGCTGCAGGGCGCGTCCA
V D A D G N V L R L E E K P D V P R S S -
CGCTCGCGCTCATCGGCGTGACGCCTTCAGCCCGGCGCTCCACGAGCGCGTACGGGCCA 2820
-----+-----+-----+-----+-----+
GCGAGCGGAGTAGCCGCACATGCGGAAGTCGGGCGGCGAGGTGCTCCGCCATGCCCGGT
L A L I G V Y A F S P A V H E A V R A I -
TCACCCCTCGCGCGCGGCGAGCTGGAGATCACCCACGCGGTGCAGTGGATGATCGACC 2880
-----+-----+-----+-----+-----+
AGTGGGGGAGGCGGGCGCGCTCGACCTCTAGTGGTGGCGGCAGGTCACTACTAGCTGG
T P S A R G E L E I T H A V Q W M I D R -
GGGGCCTGCGCGTACGGGCGGAGACCACCACCGGCCCTGGCGGACACCGGCAGCGCGG 2940
-----+-----+-----+-----+-----+
CCCCGGACGCGCATGCCCGGCTCTGGTGGTGGGCGGGACCGCGCTGTGGCGGTGCGGCC
G L R V R A E T T T R P W R D T G S A E -
AGGACATGCTGGAGGTCAACCGTCACGTCTTGGACGGACTGGAGGGCGCATCGAGGGGA 3000
-----+-----+-----+-----+-----+
TCCTGTACGACCTCCAGTTGGCAGTGCAGGACCTGCTGACCTCCCGGCGTAGCTCCCT
D M L E V N R H V L D G L E G R I E G K -
AGGTGACGCGCACAGCACGCTGCTCGGCCGGGTCCGGGTGGCCGAAGGCGCGATCGTGC 3060
-----+-----+-----+-----+-----+
TCCAGCTGCGCGTGTCTGCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
V D A H S T L V G R V R V A E G A I V R -
GGGGGTACACGTGCTGGGCGCGGCTGGTGATCGGCGCGGGTGCGGTCGTCAGCAACTCCA 3120
-----+-----+-----+-----+-----+
CCCCAGTGTGCACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
G S H V V G P V V I G A G A V V S N S S -
GTGTGCGGCGGCTACACCTCCATCGGGGAGGACTGCCGGGTGAGGACAGCGCCATCGAGT 3180
-----+-----+-----+-----+-----+
CACAGCGGGCATGTGGAGGTAGCCCCCTCCTGACGGGCGGCGGCGGCGGCGGCGGCGGCGG
V G P Y T S I G E D C R V E D S A I E Y -
ACTCCGTCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3240
-----+-----+-----+-----+-----+
TGAGGCAGGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
S V L L R G A Q V E G A S R I E A S L I -
TCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3300
-----+-----+-----+-----+-----+
AGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
G R G A V V G P A P R L P Q A H R L V I -

TCGGCGACCAACAGCAAGGTGTATCTCACCCCATGACCACGACCATCCTCGTCACCGGGGG
AGCCGCTGGTGTCTGTTCCACATAGAGTGGGGTACTGGTGCTGGTAGGAGCAGTGGCCGCC 3360

G D H S K V Y L T P * M T T T I L V T G G -
acbd

S
m
a
I

AGCGGGCTTCATTCTGCTCCGCTACGTCCGCCGGCTCCTGTGCGCCGGGGCCCCCGGGCGG
TCGCCCCGAAGTAAGCGAGGCGGATGCAGGGCGCGAGGACAGCGGGCCCCGGGGGGCGCC 3420
A G F I R S A Y V R R L L S P G A P G G -

CGTCGCGGTGACCGTCTCGACAACTCACCTACGCCGGCAGCCTCGCCCGCTGCACCG
GCAGCGCCACTGGCAGGAGCTGTTTGAAGTGGATGCGGCCGTGCGAGCGGGCGGACGTGCG 3480
V A V T V L D K L T Y A G S L A R L H A -

GGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGT
CCACGCACTGGTAGGGCCGGAGTGAAGCAGGTCCCGCTGCACACGCTGTGGCGCGAGCA 3540
V R D H P G L T F V Q G D V C D T A L V -

CGACACGCTGGCCGCGCGGCACGACGACATCGTGCCTTCGCGGCCGAGTGCACGTCGA
GCTGTGCGACCGGCGCGCGCTGCTGCTGTAGCACGTGAAGCGCCGGCTCAGCGTGCAGCT 3600
D T L A A R H D D I V H F A A E S H V D -

CCGCTCCATCACCGACAGCGGTGCCTTCACCCGCGACCAACGTGCTGGGCACCCAGGTCCT
GGCGAGGTAGTGGCTGTGCGCCACGGAAGTGGGCGTGGTTGCACGACCCGCTGGGTCCAGGA 3660
R S I T D S G A F T R T N V L G T Q V L -

GCTCGACGCGCGCTCCGCCACGGTGTGCGCACCTTCGTGCACGTCTCCACCGACGAGGT
CGAGCTGCGGCGCGAGGCGGTGCCACACGCGTGAAGCACGTGCAGAGGTGGCTGCTCCA 3720
L D A A L R H G V R T F V H V S T D E V -

GTACGGCTCCCTCCCGCACGGGGCGCGCGGAGAGCGACCCCTGCTTCCGACCTCGCC
CATGCCGAGGGAGGGCGTGCCCGGCGCGCCTCTCGCTGGGGGACGAAGGCTGGAGCGG 3780
Y G S L P H G A A A E S D P L L P T S P -

GTACGCGCGCTCGAAGGCGGCCTCGGACCTCATGGCGCTCGCCACCGACCGCACCCACGG
CATGCCGCGCAGCTTCGCGCGGAGCCTGGAGTACCGCGAGCGGGTGGTGGCGTGGGTGCC 3840
Y A A S K A A S D L M A L A H H R T H G -

CCTGGACGTCCGGGTGACCCGCTGTTTGAACAACTTCGGCCCCCACCAGCATCCCGAGAA
GGACCTGCAGGCCCCACTGGGCGACAAGCTTGTTGAAGCCGGGGGTGGTCCTAGGGCTCTT 3900
L D V R V T R C S N N F G P H Q H P E K -

GCTCATACCGCGCTTCCTGACCAGCCTCCTGTCCGGCGGCACCGTTCCTCTACGGCGA
CGAGTATGGCGCGAAGCACTGGTGGGAGGACAGGCCCGCGTGGCAAGGGGAGATGCCGCT 3960
L I P R F L T S L L S G G T V P L Y G D -

TCGGCGACCACAGCAAGGTGTATCTCACCCCATGACCACGACCATCCTCGTCACCGGGGG
AGCCGCTGGTGTCTCCACATAGAGTGGGGTACTGGTGTCTGGTAGGAGCAGTGGCCGCC 3360

G D H S K V Y L T P * M T T T I L V T G G -
acbD

S
m
a
I

AGCGGGCTTCATTGCTCCGCCTACGTCCGCCGGCTCCTGTGCCCCGGGGCCCCGGCGG
TCGCCCCGAAGTAAGCGAGGCGGATGCAGGCGGGCCGAGGACAGCGGGCCCCGGGGGCCGCC 3420
A G F I R S A Y V R R L L S P G A P G G -

CGTCGCGGTGACCGTCCCTCGACAACTCACCTACGCGGGCAGCCTCGCCCCGCTGCACCG
GCAGCGCCACTGGCAGGAGCTGTTTGAGTGGATGCGGCCGTGCGAGCGGGCGGACGTGCG 3480
V A V T V L D K L T Y A G S L A R L H A -

GGTGGCTGACCATCCCGGCCCTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGT
CCACGCACTGGTAGGGCCGGAGTGGAAGCAGGTCCCGCTGCACACGCTGTGGCGCGAGCA 3540
V R D H P G L T F V Q G D V C D T A L V -

CGACACGCTGGCCGCGCGGCACGACGACATCGTGCACTTCGCGGCCGAGTCGCACGTCGA
GCTGTGCGACCGGCGCGCCGTGCTGCTGTAGCACGTGAAGCGCCGGCTCAGCGTGCAGCT 3600
D T L A A R H D D I V H F A A E S H V D -

CCGCTCCATCACCGACAGCGGTGCCTTCACCCCGCACCAACGTGCTGGGCACCCAGGTCTT
GGCGAGGTAGTGGCTGTGCGCCACGGAAGTGGGCGTGGTTGCACGACCCGTGGGTCCAGGA 3660
R S I T D S G A F T R T N V L G T Q V L -

GCTCGACGCGCGCTCCGCCACGGTGTGCGCACCTTCGTGCACGTCTCCACCGACGAGGT
CGAGCTGCGGCGCGAGGCGGTGCCACACGCGTGAAGCACGTGCAGAGGTGGCTGCTCCA 3720
L D A A L R H G V R T F V H V S T D E V -

GTACGGCTCCCTCCCGCACGGGGCCCGCGGAGAGCGACCCCTGCTTCGACCTCGCC
CATGCCGAGGGAGGGCGTGCCCCGGCGGCGCCTCTCGCTGGGGGACGAAGGCTGGAGCGG 3780
Y G S L P H G A A A E S D P L L P T S P -

GTACGCGGCGTGAAGGCGGCCTCGGACCTCATGGCGCTCGCCACCACCGCACCCACGG
CATGCGCGCAGCTTCGCGCGGAGCCTGGAGTACCGCGAGCGGGTGGTGGCGTGGGTGCC 3840
Y A A S K A A S D L M A L A H H R T H G -

CCTGGACGTCCGGGTGACCCGCTGTTGGAACAACTTCGGCCCCCACCAGCATCCCGAGAA
GGACCTGCAGGCCCCACTGGGCGACAAGCTTGTTGAAGCCGGGGGTGGTCCTAGGGCTCTT 3900
L D V R V T R C S N N F G P H Q H P E K -

GCTCATACCGCGCTTCCTGACCAGCCTCCTGTCCGGCGGCAACGTTCCCTCTACGGCGA
CGAGTATGGCGGAAGGACTGGTCCGAGGACAGGCGCGCGTGGCAAGGGGAGATGCCGCT 3960
L I P R F L T S L L S G G T V P L Y G D -

CGGGCGGCACGTGCGCGACTGGCTGCACGTCGACGACCACGTGAGGGCCGTGGAACCTCGT
-----+-----+-----+-----+-----+ 4020
GCCCCCGGTGCACGCGCTGACCGACGTGCAGCTGCTGGTGCAGTCCCCGGCAGCTTGAGCA
G R H V R D W L H V D D H V R A V E L V -

B
g
l
I
I

CCGCGTGTGCGGGCCGGCCGGGAGAGATCTACAACATCGGGGGCGGCACCTCGCTGCCCAA
-----+-----+-----+-----+ 4080
GGCGCACAGCCCGGCCGGCCCTCTCTAGATGTTGTAGCCCCCGCGTGGAGCGACGGGTT
R V S G R P G E I Y N I G G G T S L P N -

S
s
t
I

CCTGGAGCTCACGCACCGGTTGCTCGCACTGTGCGGCGCGGGCCCGGAGCGCATCGTCCA
-----+-----+-----+-----+ 4140
GGACCTCGAGTGGCTGGCCACGAGCGTGACACGCCGCGCCCGGGCCCTCGCGTAGCAGGT
L E L T H R L L A L C G A G P E R I V H -

CGTCGAGAACC CGCAAGGGGCACGACCGGCGCTACGCGGTGACCCACAGCAAGATCACCGC
-----+-----+-----+-----+ 4200
GCAGCTCTTGCGGTTCCCGGTGCTGGCCGCGATGCGCCAGCTGGTGTCTAGTGGCG
V E N R K G H D R R Y A V D H S K I T A -

N
r
u
I

GGAACCTCGGTTACCGGCCGCGCACCGACTTCGCGACCGCGCTGGCCGACACCGCGAAGTG
-----+-----+-----+-----+ 4260
CCTTGAGCCCAATGGCCGGCGCGTGGCTGAAGCGCTGGCGCGACCGGCTGTGGCGCTTCAC
E L G Y R P R T D F A T A L A D T A K W -

GTACGAGCGGCACGAGGACTGGTGGCGTCCCTGCTCGCCGCGACATGACGTCGGGCGCG
-----+-----+-----+-----+ 4320
CATGCTCGCCGTGCTCCTGACCACCGCAGGGGACGAGCGGCGCTGTACTGCAGCCCGGCC
Y E R H E D W W R P L L A A T *

ACCGCAACCACCGGCCCGGCCGCGCACACCGCCCGCCCGCGGCGGCGGTCAG
-----+-----+-----+-----+ 4380
TGGCGTTGGTGGCCGGGGCGGCGCGTGTGGCGGCGGCGCGGCCACCGGCCGCGCAGTC
.

CGTCCGTGAGCCGGGCGCCGGCGCCCGCGGGCGGCGGCGGTGGACCCCGGACCACC
-----+-----+-----+-----+ 4440
GCAGGCACTCGGCCCGCGGCGGCGGGGCGCCCGGCCCGCCACCTGGGGGCCTGGTGG
R G H A P R R G G R P G A A T S G R V V -

E
c
o
R
I

AGTTCCGGCATGAAGACGAATTCGGTGGCGGGCGGCGGCGTTCCGCTCATCTCCTCCAGC
-----+-----+-----+-----+ 4500
TCAAGGCCGTACTTCTGCTTAAGCCACGCGCCCGCCCGCAAGGCGAGTAGAGGAGGTG
L E P M F V F E T R P P P T G S M E E L -

AGTGCCTCCACGGCGACCTGCCCATCGCCTTGACGGGCTGTCTGATGGTGGTCAGGGGA
-----+-----+-----+-----+-----+ 4560
TCACGCAGGTGCCGCTGGACGGGCTAGCGGAAGTGGCGACAGACTACCACCAGTCCCT
L A D V A V Q G M A K V P Q R I T T L P -
GGGTGGTGAAGGCCATGAGCGGCGAGTCGTGGAAGCCGACCACCGAGATGTCACCGGGA
-----+-----+-----+-----+-----+ 4620
CCCAGCCACTTCCGGTACTCGCCGCTCAGCAGCTTCGGCTGGTGGCTCTACAGTGGCCCT
P D T F A M L P S D D F G V V S I D G P -
ACCGTGAGACCCCGCCGGCGCGGCGCCCGCACGGCGCCGAGGGCCATCATGTGCTGGCG
-----+-----+-----+-----+-----+ 4680
TGGCACTCTGGGGCGGCGCGCGCGCGGCGTGGCGCGGCTCCCGGTAGTACAGCGACCGC
V T L G R R R A A R V A G L A M M D S A -
CACATGACGGCGGTGCAGCCAGGTGATCAGCGCGGACGCGGGCGGCTGGCCCCCTCC
-----+-----+-----+-----+-----+ 4740
GTGTACTGCCGCCACGTGCGGTCCAGCTAGTCGCGCCTGCGCCGCGCGGACCGGGGGAGG
C M V A T C G L D I L A S A A A Q G G E -

S
s
t
I

AGGGAGAACAGCGAGTGCTGCACGAGCTCCTCGGACTCCCGCGCCGACACTCCCAGGTGC
-----+-----+-----+-----+-----+ 4800
TCCCTCTTGTGCTCACGACGTGCTCGAGGAGCCTGAGGGCGCGGCTGTGAGGGTCCACG
L S F L S H Q V L E E S E R A S V G L H -
TCCCGCACGCCGGCGCCGGAACCCCTCGATCTTCCGCTGCACCGGCACGAAGCGGGCGGGC
-----+-----+-----+-----+-----+ 4860
AGGGCGTGCGCCGGGCTTGGGGAGCTAGAAGGCGACGTGGCCGTGCTTCGCCCGCCCG
E R V G A R F G E I K R Q V P V F R A P -
CCGACGGCGAGGCGGACGCGCTCGTGCCCCAGCTCCGCCAGGTGCGCCACGGCCAGGCGC
-----+-----+-----+-----+-----+ 4920
GGCTGCCGCTCCGGCTGCGCGAGCACGGGTGAGGCGGTCCACGCGGTGCCGGTCCGCG
G V A L G V R E H G L E A L H A V A L R -
ATCGCGGCGCGGTGCTCCGGGGAGACGAAGGTGCCTCGATCCGGGGCGAGAACCCGTTT
-----+-----+-----+-----+-----+ 4980
TAGCGCCGGGCGAGCAGGCCCCCTCTGCTTCCACGGAGCTAGGCCCGCTCTTGGGCAAG
M A A R D D P S V F P A E I R P S F G N -
ACGAGGACGAAGGGCACCTGCCGCTCGTGACCGCGCGGTACCGTCCGGTCTCGGCGGTG
-----+-----+-----+-----+-----+ 5040
TGCTCCTGCTTCCCGTGGACGGCGAGCACGTGCGCCGGCATGGCAGGCCAGAGCCGCCAC
V L V F P V Q R E H L R G Y R G T E A T -
GTGTCCGCGTGCACTCCGGAGACGAAGATGATGCCGGACACCCCGCGGTCCACGAGCATC
-----+-----+-----+-----+-----+ 5100
CACAGGCGCACGTGAGGCCTCTGCTTCTACTACGGCCTGTGGGGCGCCAGGTGCTCGTAG
T D A H L G S V F I I G S V G R D V L M -

S
m
a
I

TCCGTGAGTTCGTCTCCTCGGTGAGCCGCGCGGGTCTGCGTGGCGAGCACGGGCGTGTAG
-----+-----+-----+-----+-----+ 5160
AGGCACTCAAGCAGGAGCCAGCTCGGCGGGGCCCCAGACGCACCGCTCGTGCCCGCACATC
E T L E D E T S G G P T Q T A L V P T Y -

CCCTGACGCGTGAGCGCCTGCCCATCACCTGGGCCAGTGCGGGGAAGAAGGGGTTGTCC
-----+-----+-----+-----+-----+ 5220
GGGACTGCGCACTCGCGGACGGGGTAGTGGACCCGGTCACGCCCCCTTCTTCCCCAACAGG
G Q R T L A Q G M V Q A L A P F F P N D -
AGTTCGGGGGTGACCACTCCGACCAGCTCGGCGCGGCGCTGTGCGCGCGGCTGCTCGTAG
-----+-----+-----+-----+-----+ 5280
TCAAGCCCCCACTGGTCAGGCTGGTCGAGCCGCGCGCGACAGCGCGGCGGACGAGCATC
L E P T V L G V L E A R R Q R A P Q E Y -
CCCAGCGCGTCCAGTGCGGTGAGCACCGAGTCGCGGGGTGCCGTGGCCACACCGCGCGCA
-----+-----+-----+-----+-----+ 5340
GGGTGCGCGAGGTACGCCAGTCGTGGCTCAGCGCCACGCGCCACCGGTGTGGCGCGCGT
G L A D L A T L V S D R T G T A V G R A -
S
m
a
I
CCGTTGAGCACCCGGCTGACCGTGGCCTTGCTGACGCCCCGCGGGGCTGCGATGTGCGCG
-----+-----+-----+-----+-----+ 5400
GGCAAGTCGTGGGCGGACTGGCACCGGAACGACTGCGGGCGGGCCCGACGCTACAGCCGC
G N L V R S V T A K S V G A R A A I D A -
AGCCGCATGGTCATGGCAACGCACTCTACCTGTGCGGGGCGTCAGGGCGTGCCACCGCGC
-----+-----+-----+-----+-----+ 5460
TCGGCGTACCACTACCGTTGCGTGAGATGGACAGCCCCGCGAGTCCCGCACGGGTGGCGCG
L R M T M
acbE
GCGGAACCGCGGACTGCGGGGCACGGCCCGTCCGCCGCCACGGACCACGCGCCCGAAA
-----+-----+-----+-----+-----+ 5520
CGCCTTGGCCGCTGACGCCCGTGCCGGGCAGGCGGCGGGTGCTGCTGCGCGGGCTTT
CGATGGCTGAAAATGCTTGACGAAATGCCGCAACGTCTTTGCGCGGCTTTTCGATCCT
-----+-----+-----+-----+-----+ 5580
GCTACCGACTTTTACGAACGTCGTTTAAACGGCGTTGCAGAAAGCCGCCGAAAAGCTAGGA
GTTACGTTCTGGCAACCCCGGCGCGCGCAGAAGCGGTTGGCGTGAGGCGTCCAGACCT
-----+-----+-----+-----+-----+ 5640
CAATGCAAGGACCGTTGGGGCCGCGGCGCGTCTTCGCCAACCAGCACTCCGAGGTCTGGA
CCGCCCGATTCCGGGATCACTCAGGGGAGTTACAATGCGGCGTGGCATTGCGGCCACCG
-----+-----+-----+-----+-----+ 5700
GGCGGGCTAAGGCCCTAGTGAGTCCCTCAAGTGTTACGCCGACCGTAACGCCGGTGGC
M R R G I A A T A -
acbf
CGCTGTTGCGGGCTGTGGCCATGACGGCATCGGCGTGTGGCGGGGCGACAACGGCGGAA
-----+-----+-----+-----+-----+ 5760
GCGACAAGCGCGACACCGGTACTGCCGTAGCCGCACACCGCCCCCGCTGTTGCCGCGCT
L F A A V A M T A S A C G G G D N G S -
K
p
n
I
GCGGTACCGACGCGGGCGGCACGGAGCTGTGCGGGACCGTCACCTTCTGGGACACGTCCA
-----+-----+-----+-----+-----+ 5820
CGCCATGGCTGCGCCCGCGCTGCTCGACAGCCCTGGCAGTGGAAGACCGTGTGACGGT
G T D A G G T E L S G T V T F W D T S N -

ACGAAGCCGAGAAGGGCGACGTACCAGGCCCTCGCGGAGGGCTTCGAGAAGGAGCACCCGA
-----+-----+-----+-----+-----+ 5880
TGCTTCGGCTCTTCCGCTGCATGGTCCGGGAGCGCCTCCCGAAGCTCTTCCTCGTGGGCT
E A E K A T Y Q A L A E G F E K E H P K -
AGGTGCAGCTCAAGTACGTCAACGTCCCGTTCCGGCGAGGCGAACGCCAAGTTCAAGAACG
-----+-----+-----+-----+-----+ 5940
TCCAGCTGCAGTTTCATGCAGTTGCAGGGCAAGCCGCTCCGCTTGGGTTCAAGTTCTTGC
V D V K Y V N V P F G E A N A K F K N A -
CCGCGGGCGGCAACTCCGGTCCCCCGGACGTGATGCGTACGGAGGTGCGCTGGGTGCGCG
-----+-----+-----+-----+-----+ 6000
GGCGCCCGCGCTTGAGGCCACGGGGCCTGCACTACGCATGCCTCCAGCGGACCCAGCGCC
A G G N S G A P D V M R T E V A W V A D -
ACTTCGCCAGCATCGGCTACCTCGCCCCGCTCGACGGCACGCCCGCCCTCGACGACGGGT
-----+-----+-----+-----+-----+ 6060
TGAAGCGGTCTAGCCGATGGAGCGGGGCGAGCTGCCGTGCGGGCGGAGCTGCTCCCCA
F A S I G Y L A P L D G T P A L D D G S -
CGGACCACCTTCCCCAGGGCGGCAGCACCAGGTACGAGGGGAAGACCTACGCGGTCCCCG
-----+-----+-----+-----+-----+ 6120
GCCTGGTGGAAAGGGGTCCCGCCGTCGTGGTCCATGCTCCCCCTCTGGATGCGCCAGGGCG
D H L P Q G G S T R Y E G K T Y A V P Q -
AGGTGATCGACACCCTGGCGCTCTTCTACAACAAGGAAGTCTGACGAAGGCCGGTGTCTG
-----+-----+-----+-----+-----+ 6180
TCCACTAGCTGTGGGACCGCGAGAAGATGTTGTTCTTGACGACTGCTTCCGGCCACAGC
V I D T L A L F Y N K E L L T K A G V E -
AGGTGCCGGGCTCCCTCGCCGAGCTGAAGACGGCCGCGCCGAGATCACCGAGAAGACCG
-----+-----+-----+-----+-----+ 6240
TCCACGGCCCCGAGGGAGCGGCTCGACTTCTGCCGGCGGGGCTCTAGTGGCTCTTCTGGC
V P G S L A E L K T A A A E I T E K T G -
GCGCGAGCGGCCTCTACTGCGGGGCGACGACCCGTAAGTTGGTTCTGCCCTACCTCTACG
-----+-----+-----+-----+-----+ 6300
CGCGCTCGCCGAGATGACGCCCCGCTGCTGGGCATGAACCAAGGACGGGATGGAGATGC
A S G L Y C G A T T R T W F L P Y L Y G -
GGGAGGGCGGCGACCTGGTTCGACGAGAAGAACAAGACCGTCACGGTCGACGACGAAGCCG
-----+-----+-----+-----+-----+ 6360
CCCTCCCGCCGCTGGACCAGCTGCTCTTCTTGTCTGGCAGTGCCAGCTGCTGCTTCGGC
E G G D L V D E K N K T V T V D D E A G -
GTGTGCGCGCCTACCGCGTCATCAAGGACCTCGTGGACAGCAAGGCGGCCATCACCGACG
-----+-----+-----+-----+-----+ 6420
CACACGCGCGGATGGCGCAGTAGTTCCTGGAGCACCTGTCGTTCCGCGGGTAGTGGCTGC
V R A Y R V I K D L V D S K A A I T D A -
CGTCCGACGGCTGGAACAACATGCAGAACGCCCTTCAAGTCGGGCAAGGTGCGCATGATGG
-----+-----+-----+-----+-----+ 6480
GCAGGCTGCCGACCTTGTTGTACGTCTTGGGGAAGTTCAAGCCCGTTCCAGCGGTACTACC
S D G W N N M Q N A F K S G K V A M M V -
TCAACGGCCCCCTGGGCCATCGAGGACGTCAAGGCGGGAGCCCCGTTCAAGGACGCCGGCA
-----+-----+-----+-----+-----+ 6540
AGTTGCCGGGGACCCGGTAGCTCCTGCAGTTCCGCCCTCGGSCGAAGTTCTGCGGCCGT
N G P W A I E D V K A G A R F K D A G N -

ACCTGGGGGTGCCCCCGTCCCGGCCGGCAGTGCCGGACAGGGCTCTCCCCAGGGCGGGT
-----+-----+-----+-----+-----+ 6600
TGGACCCCGAGCGGGGGCAGGGCCGGCCGTACGGCCTGTCCCGAGAGGGGTCCCGCCCA
L G V A P V P A G S A G Q G S P Q G G W -

GGAACCTCTCGGTGTACGCGGGCTCGAAGAACCTCGACGCCTCCTACGCCTTCGTGAAGT
-----+-----+-----+-----+-----+ 6660
CCTTGGAGAGCCACATGCGCCCGAGCTTCTTGGAGCTGCGGAGGATGCGGAAGCACTTCA
N L S V Y A G S K N L D A S Y A F V K Y -

S
S
t
I

ACATGAGCTCCGCCAAGGTGCAGCAGCAGACCACCAGAGAAGCTGAGCCTGCTGCCCCACCC
-----+-----+-----+-----+-----+ 6720
TGTACTCGAGGCGGTTCCACGTCGTCTGCTGGTGGCTCTTCGACTCGGACGACGGGTGGG
M S S A K V Q Q Q T T E K L S L L P T R -

GCACGTCCGTCTACGAGGTCCCGTCCGTGCGGACAACGAGATGGTGAAGTTCTTCAAGC
-----+-----+-----+-----+-----+ 6780
CGTGCGAGCAGATGCTCCAGGGCAGGCAGCGCCTGTTGCTCTACCACTTCAAGAAGTTCG
T S V Y E V P S V A D N E M V K F F K P -

CGGCCGTGACAAGGCCGTGAAACGGCCGTGGATCGCCGAGGGCAATGCCCTCTTCGAGC
-----+-----+-----+-----+-----+ 6840
GCCGGCAGCTGTTCCGGCAGCTTGCCGGCACCTAGCGGCTCCCGTTACGGGAGAAGCTCG
A V D K A V E R P W I A E G N A L F E P -

P
S
t
I

CGATCCGGCTGCAG
-----+----- 6854
GCTAGGCCGACGTC
I R L Q -